

REMARKS

The present application was filed on September 7, 2000 with claims 1-15. In the outstanding Office Action dated November 19, 2002, the Examiner: (i) rejected claims 1-4, 6-9 and 11-14 under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,208,897 to Hutchins (hereinafter "Hutchins") in view of "Using Alternate Spellings to Generate Baseforms," IBM Technical Disclosure Bulletin, vol. 35, issue 1a, pp. 59, 1992 (hereinafter "TDB"); and (ii) indicated that claims 5, 10 and 15 are allowable.

In this response, Applicants: (i) amend claims 1-15 to correct certain minor, non-substantive informalities; and (ii) traverse the §103(a) rejections to claims 1-4, 6-9 and 11-14 for at least the reasons given below.

Applicants acknowledge allowance of claims 5, 10 and 15. Applicants have amended these claims, as well as claims 1-4, 6-9 and 11-14, to correct certain minor, non-substantive informalities.

Regarding the §103(a) rejections of claims 1-4, 6-9 and 11-14 based on the combination of Hutchins and TDB, Applicants respectfully assert that the cited combination fails to establish a prima facie case of obviousness under 35 U.S.C. §103(a), as specified in M.P.E.P. §2143.

As set forth therein, M.P.E.P. §2143 states that three requirements must be met to establish a prima facie case of obviousness. First, the cited combination must teach or suggest all the claim limitations. Second, there must be some suggestion or motivation to combine reference teachings. Third, there must be a reasonable expectation of success. While it is sufficient to show that a prima facie case of obviousness has not been established by showing that one of the requirements has not been met, Applicants respectfully believe that none of the requirements have been met.

Prior to pointing out the deficiencies in the cited combination in view of the above-described requirements, Applicants provide below a summary, illustrative explanation of the invention in a sincere effort to clarify for the Examiner what the invention is directed toward and where it may be used. Applicants believe that, given this explanation and the discussion of the deficiencies of the cited combination to follow, it will be evident that the §103(a) rejections should be withdrawn.

The present invention, for example as recited in claim 1, provides a voice information registration method, employed by a speech recognition apparatus, which comprises the steps of: (a)

obtaining a sentence group, which includes a first to an N-th sentence, wherein N is a number equal to or greater than two; (b) obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N; (c) obtaining a base form based on said sounds-like spelling of said word; and (d) registering said base form in a speech recognition dictionary in correlation with said word.

By way of further explanation, the present specification, at page 1, lines 8-19, describes one problem that the claimed invention addresses:

As is described in Japanese Unexamined Patent Publication No. Hei 10-320168, the disclosure of which is incorporated by reference herein, a conventional method is available whereby voice is used to specify information displayed on a screen. However, to use this method, a menu or a button in an application, and a sentence in which a link to a web is included must be registered using words that can be recognized by a speech recognition system.

All of the character strings for a menu, in this case, can be statically added to a speech recognition dictionary, but since the web link would tend to be changed daily, coping with such a change would exceed the capabilities of a method for which static registration is employed. In addition, if too many words, more than are necessary, are added to the dictionary, other problems, such as a reduction in the recognition accuracy or an extended processing time, may be encountered.

The present specification, at page 2, lines 4-12, then provides an illustrative description of how the registration method of the invention may be realized, thus overcoming the above-described problem:

A group of sentences to be recognized is obtained from an application, and using parsing logic, each target sentence to be recognized is divided into words, speech recognition units. Thereafter, the words in each target sentence are examined to determine whether among them there are unknown words that are not registered in the speech recognition dictionary, but for which the sounds-like spelling is available. If an unknown word is found, a base form, for which the pronunciation is inferred from the sounds-like spelling, is prepared and is registered in the speech recognition dictionary. This base form is employed when the voice of a user is recognized who has orally designated one of the sentences.

Thus, the invention is directed toward techniques for registering unknown words such that these words may then be used in techniques for recognizing speech uttered by a user. By way of one example, the unknown words may be words associated with a web link, which, as mentioned above, tend to change.

On the other hand, Hutchins discloses only techniques for recognizing speech, not registration techniques, as in the claimed invention. More particularly, as disclosed in the abstract, Hutchins is directed toward a method for speech recognition which includes steps of sampling a speaker's speech and providing speech data sample segments of predetermined length at predetermined sampling intervals based on changes in energy in the speech. Cohesive speech segments, which correspond to intervals of stable vocoids, changing vocoids, frication, and silence, are identified from the speech data sample segments, and are assigned frames of subsyllables. Each cohesive segment corresponds to at least one respective frame, and each frame includes at least one of a plurality of subsyllables that characterizes predetermined gross and fine phonetic attributes of the respective cohesive segment. The subsyllables are located in a first lookup table mapping sequences of subsyllables into syllables, and the syllables are combined into words by locating words in another lookup table. The conformance of sequences of the words to a set of predetermined checking rules is checked, and a recognition result is reported.

TDB discloses a technique for building word models for a speech recognition system that includes a user entering a "sound-like spelling."

First, with respect to independent claims 1, 2, 6, 7, 11 and 12, the combination of Hutchins and TDB fails to teach or suggest all of the limitations of such claims. For at least this reason, a prima facie case of obviousness has not been established.

The Office Action (at paragraph 5) contends that Hutchins discloses the steps/operations of independent claims 1, 2, 6, 7, 11 and 12 including: "obtaining a sentence group . . . sentence;" "obtaining a spelling . . . speech recognition dictionary;" "obtaining a base form . . . word;" and "registering said base form . . . said word." The Office Action acknowledges that Hutchins does not disclose "sound-like spellings," however, points out that TDB uses "sound-like spellings." The

Office Action then summarily concludes that it would have been obvious to combine the two references to achieve the claimed invention. Applicants strongly disagree.

Applicants do not assert that they have developed the concept of “sounds-like spellings.” Such spellings are known, for example, as evidenced by TDB. However, the automated use of “sounds-like spellings” as recited in the registration techniques of the claimed invention was not known prior to the invention and is clearly not taught or suggested by the combination of Hutchins and TDB.

As explained above, Hutchins is directed only toward speech recognition, not unknown word registration. While registered words are eventually used to recognize speech, the techniques for realizing the two concepts are significantly different. Thus, Hutchins does not disclose any registering step. Since this is the case, it is clear that Hutchins also does not disclose the steps that precede the registering step, namely, obtaining a target sentence group, obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, and obtaining a base form based on said sounds-like spelling of said word, as in the claimed invention. All the sections of Hutchins cited by the Examiner deal only with a subsyllable-based recognition system.

While TDB discusses the use of “sounds-like spellings,” TDB explains that a user may enter a “sounds-like spelling” to assist in building word models. This is significantly different than the steps/operations of the claimed invention since TDB, other than stating that a user may enter a “sounds-like spelling,” is silent as how a word model is formed given the “sounds-like spelling.” Thus, TDB suggests nothing about obtaining a sentence group, obtaining a sounds-like spelling for a word that is included in an i-th sentence, but is not entered in a speech recognition dictionary, obtaining a base form based on said sounds-like spelling of said word, and registering said base form in a speech recognition dictionary in correlation with said word, as recited in the claimed invention.

Also, with specific regard to claims 2, 7 and 12, despite a contention in the Office Action to the contrary, neither Hutchins nor TDB disclose obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to the i-th sentence, as in the claimed invention.

Therefore, as pointed out above, since each of Hutchins and TDB do not teach or suggest limitations of claims 1, 2, 6, 7, 11 and 12, their combination fails to do so.

Second, with respect to independent claims 1, 2, 6, 7, 11 and 12, Applicants assert that no motivation or suggestion exists to combine Hutchins and TDB. For at least this reason, a prima facie case of obviousness has not been established. Hutchins discloses a subsyllable-based speech recognition system, while TDB discloses a word model building technique that uses “sounds-like spellings.” Applicants fail to see the motivation or suggestion to combine the very specific subsyllable speech recognition techniques of Hutchins with the word model building techniques of TDB. While both references generally relate to aspects of speech recognition, Applicants strongly believe that one ordinarily skilled in the art would not look to the word model building techniques of TDB to find inspiration to improve the very specific subsyllable speech recognition techniques of Hutchins, or visa versa.

Lastly, with respect to independent claims 1, 2, 6, 7, 11 and 12, Applicants assert that there is no reasonable expectation of success in achieving the present invention through a combination of Hutchins and TDB. For at least this reason, a prima facie case of obviousness has not been established. Despite the assertion in the Office Action, Applicants do not believe that Hutchins and TDB are combinable since it is not clear how one would combine them. There is no guidance provided in the present Office Action. However, even if combined, for the sake of argument, they would not achieve the automated registration techniques of the claimed invention.

Therefore, for at least the reasons given above, Applicants respectfully request that the §103(a) rejections of independent claims 1, 2, 6, 7, 11 and 12 be withdrawn.

Furthermore, it is respectfully asserted that the claims which respectively depend from independent claims 2, 7 and 12, i.e., claims 3, 4, 8, 9, 13 and 14, are patentable over the cited combinations. Thus, Applicants request withdrawal of the §103(a) rejections of said claims.

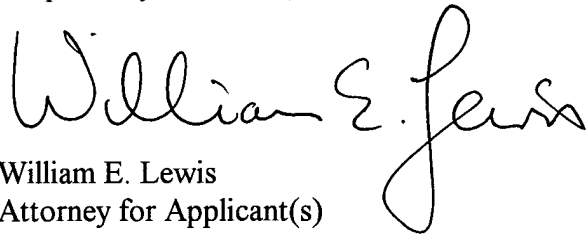
In addition, it is asserted that dependent claims 3, 4, 8, 9, 13 and 14 recite patentable subject matter in their own rights. By way of example, claims 3, 8 and 13 recite that the group of sentences is obtained from an application and that a control message corresponding to the i-th sentence is generated and transmitted to the application. There is no application or control message disclosed

in Hutchins or TDB. Further, claims 4, 9 and 14 recite that a sounds-like spelling score is stored in correlation with the sounds-like spelling of the word, and that a pronunciation score is stored in correlation with the base form, and that the base form is registered in a speech recognition dictionary when a function value that is obtained by using the sounds-like spelling score and the pronunciation score exceeds a threshold value. Again, Hutchins and TDB are completely silent as to the use of any such scores.

Attached hereto is a marked-up version of the changes made to the specification and claims by the present Amendment.

In view of the above, Applicants believe that claims 1-15 are in condition for allowance, and respectfully request withdrawal of the §103(a) rejection.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "William E. Lewis". The signature is fluid and cursive, with the first name "William" being the most prominent part.

Date: February 19, 2003

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946



VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE SPECIFICATION

The paragraph beginning at page 26, line 13, has been amended as follows:

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be [affected therein] made by one skilled in the art without departing from the scope or spirit of the invention.

IN THE ABSTRACT

The abstract has been amended as follows:

A group of sentences to be recognized is obtained from an application[,] and, using parsing logic, each target sentence to be recognized is divided into words, e.g. speech recognition units. Thereafter, the words in each target sentence are examined to determine whether among them there are unknown words that are not registered in the speech recognition dictionary, but for which the sounds-like spelling is available. If an unknown word is found, a base form, for which the pronunciation is inferred from the sounds-like spelling, is prepared and is registered in the speech recognition dictionary. This base form is employed when the voice of a user, [is recognized] who has orally designated one of the sentences, is recognized.

IN THE CLAIMS

Claims 1-15 have been amended as follows:

1. (Amended) A voice information registration method, employed by a speech recognition apparatus[, for which a voice input device is used], comprising:

(a) obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

- (c) obtaining a base form based on said sounds-like spelling of said word; and
- (d) registering said base form in a speech recognition dictionary in correlation with said word.

2. (Amended) A sentence specification method, employed by a speech recognition apparatus[, for which a voice input device is used], comprising:

a registration step including:

(a1) obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two,

(a2) obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N,

(a3) obtaining a base form based on said sounds-like spelling of said word, and

(a4) registering said base form in a speech recognition dictionary in correlation with said word; and

a recognition step including:

(b1) obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence,

(b2) employing said base form to recognize said voice information and to select a speech recognition sentence, and

(b3) comparing said i-th sentence with said selected speech recognition sentence.

3. (Amended) The sentence specification method according to claim 2, wherein said group of [target] sentences is obtained from an application, said method further comprising a step of generating a control message corresponding to said i-th sentence and transmitting said control message to said application.

4. (Amended) The sentence specification method according to claim 2, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word[;], wherein a

pronunciation score is stored in correlation with said base form[;], and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said base form is registered in a speech recognition dictionary.

5. (Amended) A sentence specification method, employed by a speech recognition apparatus[, for which a voice input device is used], comprising:

a registration step including:

(a1) obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two,

(a2) obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N,

(a3) obtaining a base form based on said sounds-like spelling of said word,

(a4) calculating a score for said base form, and

(a5) registering said base form, when said score for said base form exceeds a threshold value, in said speech recognition dictionary in correlation with said word; and

a recognition step including:

(b1) obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence,

(b2) employing said base form to recognize said voice information and to select a speech recognition sentence,

(b3) comparing said i-th sentence with said selected speech recognition sentence,

(b4) performing a process associated with a sentence for which a match is obtained when it is found that said i-th sentence and said selected speech recognition sentence match,

(b5) updating said threshold value to provide a smaller second threshold value when only part of said selected speech recognition sentence is matched by said i-th sentence,

(b6) detecting an unknown word that is included in said i-th sentence,

(b7) obtaining [the] a sounds-like spelling of said unknown word,

- (b8) obtaining a second base form based on said sounds-like spelling for said unknown word,
- (b9) calculating the score for said second base form, and
- (b10) registering said score in said speech recognition dictionary, in correlation with said unknown word, when said score for said second base form exceeds said second threshold value.

6. (Amended) A speech recognition apparatus[, for which a voice input device is used], comprising:

(a) a sentence specification unit for obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two[:],]

(b) an unknown word detector for obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N[:],]

(c) a base form generator for obtaining a base form based on said sounds-like spelling of said word; and

(d) a speech recognition dictionary to which said base form is stored in correlation with said word.

7. (Amended) A speech recognition apparatus[, for which a voice input device is used], comprising:

(a) a sentence specification unit for obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) an unknown word detector for obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) a base form generator for obtaining a base form based on said sounds-like spelling of said word;

(d) a speech recognition dictionary in which said base form is stored in correlation with said word;

(e) a voice input unit for obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence; and

(f) a speech recognition engine for employing said base form to recognize said voice information and to select a speech recognition sentence[,];

wherein [(a2)] said sentence specification unit compares said i-th sentence with said selected speech recognition sentence.

8. (Amended) The speech recognition apparatus according to claim 7, wherein said sentence specification unit obtains said group of [target] sentences from an application, generates a control message corresponding to said i-th sentence, and transmits said control message to said application.

9. (Amended) The speech recognition apparatus according to claim 7, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word[,], wherein a pronunciation score is stored in correlation with said base form[,], and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said base form is registered in a speech recognition dictionary.

10. (Amended) A speech recognition apparatus[, for which a voice input device is used], comprising:

(a) a sentence specification unit for obtaining a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) an unknown word detector for obtaining [the] a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) a base form generator for obtaining a base form based on said sounds-like spelling of said word, and for calculating a score for said base form;

(d) a speech recognition dictionary in which, when said score for said base form exceeds a threshold value, said base form is registered in said speech recognition dictionary in correlation with said word;

(e) a voice input unit for obtaining voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence; and

(f) a speech recognition engine for employing said base form to recognize said voice information and to select a speech recognition sentence[.];

wherein [(a2)] said sentence specification unit compares said i-th sentence with said selected speech recognition sentence[.], performs a process associated with a sentence for which a match is obtained when it is found that said i-th sentence and said selected speech recognition sentence match[.], updates said threshold value to provide a smaller second threshold value when only part of said selected speech recognition sentence is matched by said i-th sentence[.], instructs said unknown word detector to detect an unknown word that is included in said i-th sentence and to obtain [the] a sounds-like spelling of said unknown word[.], instructs said base form generator to obtain a second base form based on said sounds-like spelling for said unknown word and to calculate the score for said second base form[.], and registers said score in said speech recognition dictionary, in correlation with said unknown word, when said score for said second base form exceeds said second threshold value.

11. (Amended) A storage medium in which a program for specifying a sentence is stored to be executed by a speech recognition apparatus[, for which a voice input device is used], said program comprising:

(a) program code for instructing said speech recognition apparatus to obtain a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) program code for instructing said speech recognition apparatus to obtain a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) program code for instructing said speech recognition apparatus to obtain a base form based on said sounds-like spelling of said word; and

(d) program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary in correlation with said word.

12. (Amended) A storage medium in which a program for specifying a sentence is stored to be executed by a speech recognition apparatus[, for which a voice input device is used], said program comprising:

(a) program code for instructing said speech recognition apparatus to obtain a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) program code for instructing said speech recognition apparatus to obtain a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) program code for instructing said speech recognition apparatus to obtain a base form based on said sounds-like spelling of said word;

(d) program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary in correlation with said word;

(e) program code for instructing said speech recognition apparatus to obtain voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence;

(f) program code for instructing said speech recognition apparatus to employ said base form to recognize said voice information and to select a speech recognition sentence; and

(g) program code for instructing said speech recognition apparatus to compare said i-th sentence with said selected speech recognition sentence.

13. (Amended) The storage medium according to claim 12, wherein said group of [target] sentences is obtained from an application, and wherein program code is stored to instruct said speech recognition apparatus to generate a control message corresponding to said i-th sentence and to transmit said control message to said application.

14. (Amended) The storage medium according to claim 12, wherein a sounds-like spelling score is stored in correlation with the sounds-like spelling of said word[;], wherein a pronunciation score is stored in correlation with said base form[;], and wherein, when a function value that is obtained by using said sounds-like spelling score and said pronunciation score exceeds a threshold value, said base form is registered in a speech recognition dictionary.

15. (Amended) A storage medium in which a program for specifying a sentence is stored to be executed by a speech recognition apparatus[, for which a voice input device is used], said program comprising:

(a) program code for instructing said speech recognition apparatus to obtain a sentence group, which includes [the] a first to [the] an N-th [(N is a natural number equal to or greater than 2)] sentence, wherein N is a number equal to or greater than two;

(b) program code for instructing said speech recognition apparatus to obtain a sounds-like spelling for a word that is included in [the] an i-th [(i is a natural number equal to or smaller than N)] sentence, but is not entered in a speech recognition dictionary, wherein i is a number equal to or less than N;

(c) program code for instructing said speech recognition apparatus to obtain a base form based on said sounds-like spelling of said word;

(d) program code for instructing said speech recognition apparatus to calculate a score for said base form;

(e) program code for instructing said speech recognition apparatus to register said base form, when said score for said base form exceeds a threshold value, in said speech recognition dictionary in correlation with said word;

(f) program code for instructing said speech recognition apparatus to obtain voice information that is input as a user reads and vocally reproduces a display corresponding to said i-th sentence;

(g) program code for instructing said speech recognition apparatus to employ said base form to recognize said voice information and to select a speech recognition sentence;

(h) program code for instructing said speech recognition apparatus to compare said i-th sentence with said selected speech recognition sentence;

(i) program code for instructing said speech recognition apparatus to perform a process associated with a sentence for which a match is obtained when it is found that said i-th sentence and said selected speech recognition sentence match;

(j) program code for instructing said speech recognition apparatus to update said threshold value to provide a smaller second threshold value when only part of said selected speech recognition sentence is matched by said i-th sentence;

(k) program code for instructing said speech recognition apparatus to detect an unknown word that is included in said i-th sentence,

(l) program code for instructing said speech recognition apparatus to obtain [the] a sounds-like spelling of said unknown word;

(m) program code for instructing said speech recognition apparatus to obtain a second base form based on said sounds-like spelling for said unknown word;

(n) program code for instructing said speech recognition apparatus to calculate the score for said second base form; and

(o) program code for instructing said speech recognition apparatus to register[ing] said score in said speech recognition dictionary, in correlation with said unknown word, when said score for said second base form exceeds said second threshold value.